Alzheimer's Research & Therapy

Prediction of amyloid pathology in cognitively unimpaired individuals using voxelwise analysis of longitudinal structural brain MRI

--Manuscript Draft--

Manuscript Number:	
Full Title:	Prediction of amyloid pathology in cognitively unimpaired individuals using voxelwise analysis of longitudinal structural brain MRI
Article Type:	Research
Funding Information:	Ministerio de Economía, Industria y Competitividad, Gobierno de España (MALEGRA TEC2016-75976-R) — Dr Verónica Vilaplana European Regional Development Fund (MALEGRA TEC2016-75976-R) — Dr Verónica Vilaplana Ministerio de Educación, CUltura y Deporte (FPU) — Mr Adrià Casamitjana Ministerio de Economía, Industria y Competitividad, Gobierno de España (RYC-2013-13054) — Dr Juan Domingo Gispert
Abstract:	Background Magnetic resonance imaging (MRI) has unveiled specific alterations at different stages of Alzheimer's disease (AD) pathophysiologic continuum constituting what has been established as 'AD signature'. To what extent MRI can detect amyloid-related cerebral changes from structural MRI in cognitively unimpaired individuals is still an area open for exploration. Method Longitudinal 3D-T1 MRI scans were acquired from a subset of the ADNI cohort comprising 403 subjects: 79 controls (Ctrls), 50 preclinical AD (PreAD), 274 MCI and dementia due to AD (MCI/AD). Amyloid CSF was used as gold-standard measure with established cut-offs (<192pg/mL) to establish diagnostic categories. Cognitively unimpaired individuals were defined as Ctrls if were amyloid negative and PreAD otherwise. The MCI/AD group was amyloid positive. Only subjects with the same diagnostic category at baseline and follow-up visits were considered for the study. Longitudinal morphometric analysis was performed using SPM12 to calculate Jacobian determinant maps. Statistical analysis was carried out on these jacobian maps to identify structural changes that were significantly different between diagnostic categories. A machine learning classifier was applied on Jacobian determinant maps to predict the presence of abnormal amyloid levels in cognitively unimpaired individuals. The performance of this classifier was evaluated using receiver operating characteristic curve analysis and as a function of the follow-up time between MRI scans. We applied a cost function to assess the benefit of using this classifier in the triaging of individuals in a clinical trial-recruitment setting. Results The optimal follow-up time for classification of Ctrls vs PreAD was $\Delta t>2.5$ years. The longitudinal voxel-based classifier achieved an AUC=0.866 (95%CI:0.72-0.97). The brain regions that showed the highest discriminative power to detect amyloid abnormalities were the medial, inferior and lateral temporal lobes, precuneus, caudate heads, basal forebrain and lateral ventricles. Conclusions Our work supports that machine learning applied to longitudinal brain volumetric changes can be used to predict, with high precision, presence of amyloid abnormalities in cognitively unimpaired subjects. Used as a triaging method to identify a fixed number of amyloid positive individuals, this longitudinal voxelwise classifier is expected

	to avoid 55% of unnecessary CSF and/or PET scans and reduce economic cost by 40%.
Corresponding Author:	Juan Domingo Gispert, PhD BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation Barcelona, SPAIN
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation
Corresponding Author's Secondary Institution:	
First Author:	Paula M Petrone
First Author Secondary Information:	
Order of Authors:	Paula M Petrone
	Adrià Casamitjana
	Carles Falcon
	Miquel Artigues
	Grégory Operto
	Raffaele Cacciaglia
	José Luis Molinuevo
	Verónica Vilaplana
	Juan Domingo Gispert, PhD
Order of Authors Secondary Information:	
Additional Information:	
Question	**Response**
Is this study a clinical trial?<hr>A clinical trial is defined by the Word Health Organisation as 'any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes'.	No

Prediction of amyloid pathology in cognitively unimpaired individuals using voxelwise analysis of longitudinal structural brain MRI

Paula M. Petrone[‡1], Adrià Casamitjana[‡2], Carles Falcon[1,3], Miquel Artigues[2], Grégory Operto[1], Raffaele Cacciaglia[1], José Luis Molinuevo[1,4,5], Verónica Vilaplana[2*], Juan Domingo Gispert[1,4,5*] for the Alzheimer's Disease Neuroimaging Initiative[$]

[1] Barcelonaβeta Brain Research Center (BBRC), Pasqual Maragall Foundation. Barcelona, Spain.

[2] Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, Barcelona, Spain

[3] Centro de Investigación Biomédica en Red Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN). Spain

[4] CIBER Fragilidad y Envejecimiento Saludable (CIBERFES), Madrid, Spain

[5] Universitat Pompeu Fabra, Barcelona, Spain

[‡]Authors contributed equally.

[$]Data used in precreation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at (http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

* Correspondence to:

Dr. Juan Domingo Gispert

Barcelonaβeta Brain Research Centre - Pasqual Maragall Foundation

C/ Wellington 30. 08005 Barcelona, Spain

Email: jdgispert@barcelonabeta.org

Dr. Veronica Vilaplana

Department of Signal Theory and Communications, Universitat Politècnica de Catalunya

C/ Jordi Girona 1-3, edifici D5 Campus Nord UPC. 08034 Barcelona, Spain

Email: veronica.vilaplana@upc.edu

Co-author email addresses:

PMP, petronepau@gmail.com; AC, adria.casamitjana@upc.edu; CF, cfalcon@barcelonabeta.org; MA, miquelartigues1995@gmail.com; GO, goperto@barcelonabeta.org; RC, rcacciaglia@barcelonabeta.org; JLM, jlmolinuevo@barcelonabeta.org; VV, veronica.vilaplana@upc.edu; JDG, jdgispert@barcelonabeta.org

Abstract

Background: Magnetic resonance imaging (MRI) has unveiled specific alterations at different stages of Alzheimer's disease (AD) pathophysiologic continuum constituting what has been established as 'AD signature'. To what extent MRI can detect amyloid-related cerebral changes from structural MRI in cognitively unimpaired individuals is still an area open for exploration.

Method: Longitudinal 3D-T1 MRI scans were acquired from a subset of the ADNI cohort comprising 403 subjects: 79 controls (Ctrls), 50 preclinical AD (PreAD), 274 MCI and dementia due to AD (MCI/AD). Amyloid CSF was used as gold-standard measure with established cut-offs (<192pg/mL) to establish diagnostic categories. Cognitively unimpaired individuals were defined as Ctrls if were amyloid negative and PreAD otherwise. The MCI/AD group was amyloid positive. Only subjects with the same diagnostic category at baseline and follow-up visits were considered for the study. Longitudinal morphometric analysis was performed using SPM12 to calculate Jacobian determinant maps. Statistical analysis was carried out on these jacobian maps to identify structural changes that were significantly different between diagnostic categories. A machine learning classifier was applied on Jacobian determinant maps to predict the presence of abnormal amyloid levels in cognitively unimpaired individuals. The performance of this classifier was evaluated using receiver operating characteristic curve analysis and as a function of the follow-up time between MRI scans. We applied a cost function to assess the benefit of using this classifier in the triaging of individuals in a clinical trial-recruitment setting.

Results: The optimal follow-up time for classification of Ctrls vs PreAD was $\Delta t>2.5$ years. The longitudinal voxel-based classifier achieved an AUC=0.866 (95%CI:0.72-0.97). The brain regions that showed the highest discriminative power to detect amyloid abnormalities were the medial, inferior and lateral temporal lobes, precuneus, caudate heads, basal forebrain and lateral ventricles.

Conclusions: Our work supports that machine learning applied to longitudinal brain volumetric changes can be used to predict, with high precision, presence of amyloid abnormalities in cognitively unimpaired subjects. Used as a triaging method to identify a fixed number of amyloid positive individuals, this

longitudinal voxelwise classifier is expected to avoid 55% of unnecessary CSF and/or PET scans and reduce economic cost by 40%.

Keywords: preclinical AD signature, Jacobian determinant, MRI, machine learning, longitudinal voxelwise analysis

1. Background

Despite enormous efforts, there is yet no disease-modifying treatment available for Alzheimer's disease (AD). In this scenario, a promising strategy aims to prevent AD by developing interventions before the onset of symptoms [1]. The main challenge to operationalize such strategy lies in the detection of those individuals who are at increased risk to develop symptoms in the short term and would best benefit from these interventions [2].

Biomarker studies have demonstrated that AD pathology unfolds as a continuum [3]. AD starts with a dormant asymptomatic stage - the "preclinical state" (PreAD) - followed by the progressively impaired symptomatic states of mild cognitive impairment (MCI) and dementia. PreAD is characterized by unimpaired cognition, performance within norms taking into account age and education, and abnormal amyloid biomarkers as measured in cerebrospinal fluid (CSF) or by positron emission tomography (PET). The PreAD stage can last for decades and thus provides a window of opportunity for potential preventive intervention with disease-modifying therapies as long as the earliest pathophysiological changes that precede the emergence of AD clinical symptoms can be detected. However, CSF and PET are not suitable techniques for the screening or triaging of the general population given their invasiveness and high cost.

Recent developments in Magnetic Resonance Imaging (MRI) permit the study of the neuroanatomy with unprecedented detail. MRI has proven to be instrumental at characterizing impending dementia and cognitive decline due to AD both for research and in the clinic [4]. The neuroimaging AD signature has been established as structural changes in AD-vulnerable structures (i.e. entorhinal cortex, hippocampus and temporal lobe) that constitute diagnostic markers of cognitive impairment and AD progression [5, 6]. A *preclinical AD signature* might also be present in structural imaging, as several recent studies point out [7, 8, 9, 10, 13], even though to a lower degree as what is observed in the clinical stages of the disease. On top of this, preliminary results by our group [13] and others [11] show that brain anatomical changes at the PreAD stage involve regions of the aforementioned AD signature.

In this line, artificial intelligence, hand in hand with MRI, come to the aid of early disease detection across a variety of medical domains. In the scope of AD, many efforts have been dedicated to the automated

detection of mild-cognitive impairment and dementia due to AD based on biomarkers and MRI-T1 images of subjects [12]. However, the detection of PreAD from MRI datasets has received much less attention. In a previous study, we showed that MRI in combination with machine learning can predict amyloid positivity with enough accuracy to be cost-effective as a pre-screening tool [13]. In that report, the MRI's predictive capacity was validated in two independent cohorts and a similar cross-sectional study achieved similar results in a third population [14]. In the present work, instead, we investigate how pathological brain structural changes in preAD and AD subjects differ from normal brain ageing processes. Our longitudinal voxelwise approach utilizes tensor-based morphometry to make inferences on local tissue gain or loss that occur over the different stages of AD. We seek to identify the most significant features from the Jacobian determinant maps that are characteristic of AD progression, and use these to predict amyloid positivity in early AD stages.

The objectives of this work are therefore twofold. On one hand, we seek to characterize the PreAD signature based on a statistical analysis of Jacobian determinants across the sample population. Secondly, we develop a machine-learning tool to identify amyloid positive subjects through the detection of abnormal structural changes in their brain. This novel classification model relies on longitudinal MRI images acquired throughout two time points, and is able to predict amyloid positivity based solely on brain structural changes that are different to those that pertain to normal brain ageing as shown in cognitively unimpaired and amyloid negative individuals used as controls.

2. Methods

2.1 Subjects

Subjects for this study were selected from the ADNI database [15] provided that they had two or more longitudinal 3D-T1 MRI acquisitions and cerebrospinal fluid (CSF) biomarker data publicly available. Subjects were assigned biomarker-assisted diagnostic categories following recently published guidelines [16]. Subjects labelled as 'Normal' in ADNI were classified as amyloid negative cognitively unimpaired (Ctrl) if CSF Aβ was above 192 pg/mL and preclinical (PreAD) if CSF Aβ was below 192 pg/mL. This threshold has been shown to optimally discriminate between cognitively unimpaired individuals and AD patients, and has been extensively used as cutoff value for amyloid positivity [17]. Subjects were categorized as MCI or AD according to the ADNI diagnostic categories reported in [18] and we selected only those individuals with CSF Aβ levels below 192 pg/mL to exclude subjects harboring non-AD pathological changes. At baseline, this diagnostic algorithm yielded 79 Ctrl, 50 PreAD and 274 MCI/dementia due to AD, a total of 403 subjects with complete imaging and CSF data. As an additional inclusion criteria, in follow-up visits, all subjects remain in the same diagnostic category.

2.2 MRI data

Structural 3D-T1 MRI images were acquired across different scanners and institutions. Each image was associated with a cognition score and a set of CSF biomarker values (amyloid-beta, total tau and phosphorylated tau). The date of the CSF extraction was selected to be within 90 days from the date of the MRI scan. Each subject had at least one follow-up visit with the corresponding T1-MRI image, cognition score and CSF biomarker values. The number of visits may differ across subjects (Table 1). The total number of MRI scans analyzed was 980. The time interval between visits was, at least, 6 months apart.

2.2.1 Image Analysis

The SPM12 [19] neuroimaging software suit was used for every step of this longitudinal analysis pipeline. All image pairs corresponding to the same subject from the ADNI database were processed with longitudinal pairwise registration. Images in each pair were averaged and their respective Jacobian determinant was calculated, which reflects the regional cerebral volumetric changes between the

respective time-points. DARTEL normalization [33] was applied onto average images to normalize Jacobian determinant maps to MNI space [34] and allow comparison across subjects. The intensity of each voxel in the Jacobian image was normalized by the interval of time between reference and follow-up visits (i.e. Δt). The number of Jacobian determinant maps for each subject's diagnostic category is: 184 Ctrl, 114 PreAD and 543 MCI/AD.

On top of the voxelwise analysis, a regional analysis was also performed. To this end, regions of interest (ROIs) in the AAL atlas were masked by each subject's grey matter segmentation and the mean value of the remaining voxels' intensity per region was computed [20].

2.3 Automated recognition of PreAD volumetric changes using machine learning

All Jacobian determinant maps from each subject were labeled using subject's label (i.e. PreAD, Ctrl), leaving a study cohort of N = 129 (N_{Ctrl} = 79, N_{PreAD} = 50). Importantly, as mentioned before, we only consider pairs of images for which no transitions have been observed across categories. This analysis was performed only on the PreAD and Ctrl subjects.

Feature selection

Due to the limited sample size and high-dimensionality of the Jacobian determinant maps, we perform feature selection to keep an optimal percentage of the most relevant features. To this end, we use a filter feature selection method based on F-test, taking into consideration Jacobian features and subject labels. The F-test metric is used to create a ranking of all Jacobian features and finally a fixed percentage of the highly ranked features are used for classification [22].

Classification and performance evaluation

Ridge logistic regression with hyperparameter C [23] is used for binary classification of Jacobian features within the nested cross-validation (CV) framework [24] defined in Figure 1. It consists of an inner CV loop for model selection and an outer CV loop for assessing model performance. First, in the outer loop, subjects are randomly divided into 80% train set and 20% test set previously fixing a prevalence of interest (the percentage of samples of the amyloid positive class). For each subject in either set, all available Jacobian determinant maps are used for classification. The train set is used for feature selection

and model optimization while the test set is left out for final model evaluation. The random split by subject ensures that there is no contamination of the test set with Jacobian determinants of the train set.

Feature selection is computed using only the train set. In the model optimization step, the train set is further split into sub-train (2/3) and validation (1/3) sets using a (k=3)-fold cross-validation. A grid search strategy is used to optimize the classifier hyperparameter C by maximizing the f1-score on the validation set. Finally, the model is estimated using the optimized hyperparameter C on the whole train set. Then, the model is applied to the test set to compute standard performance metrics (i.e. area under the receiver operating curve (AUC), accuracy, precision, sensitivity, specificity and f1-score). Following the formulation in [13], we also report the reduction of economic cost (i.e. savings) of using this classification framework as a tool for AD screening.

This procedure is repeated n=100 times and performance results are reported using the average and standard deviation. The overall implementation is based on the scikit-learn Python library, (version 0.18) [25].

Savings

Savings in a triaging process were calculated as the percentage difference in resources between the standard recruitment protocol and using our proposed protocol in [13] to obtain a desired number of PreAD subjects for the clinical study. Savings were assessed in terms of economic cost [Eq. 1] or participant burden [Eq. 2], i.e. the amount of unnecessary PET/CSF tests spared through MRI-screening.

Eq. [1] $$Savings_{COST} = 1 - \frac{1}{2 \cdot C_{avg}} \left(\rho \frac{C_{PET}}{P} + \frac{C_{MRI}}{R} \right)$$

Eq. [2] $$Savings_{CSF/PET} = 1 - \rho \frac{1}{P}$$

Savings rely on the algorithmic precision (P) and recall/sensitivity (R) and on the prevalence of the population (ρ). The costs of MRI and PET were estimated as C_{MRI} = 700€, and C_{CSF} = 3000€ and C_{avg} represents the average cost among the screening tests which may include additional costs (e.g. neuropsychological cognitive testing).

2.4 Statistical Analysis

The aim of the statistical analysis is to identify significant group differences in brain volumetric change rate between AD stages. We will investigate the location of these stage-specific changes, and whether they represent a volume increase (positive changes) or decrease (negative changes). Every Jacobian determinant map is treated as an independent variable.

Two-sample t-test: Statistical analyses were performed by comparing any combination of two subject categories. The uncorrected threshold for statistical significance was $p < 0.005$. Spatial clustering of regions with statistically relevant voxels was applied to rule out false-positives, with a clustering threshold of $k > 100$ voxels under which voxel clusters with smaller sizes were discarded.

Data normalization: The effects of normal ageing on brain structural changes was considered as a confounder and regressed out [21]. Coefficients for linear regression on age were fitted using only Ctrls (i.e. individuals that are amyloid negative, asymptomatic in all visits).

The age corresponding to each Jacobian determinant was defined as the mean age between the two visits i.e. $age = (age_{reference} + age_{follow\text{-}up})/2$.

3. Results

3.1.- Demographic and follow-up comparisons

We included a total of 430 subjects at baseline with at least one follow-up visit over three categories: Ctrl (n=79), PreAD (n=50) and MCI/AD (n=274). Demographic data and follow-up period is presented in Table 2 split into different categories.

We denote as Δt the time interval between two follow-up visits (i.e. reference and target images).

The distribution of the time interval (*Δt*) between follow-up visits on all subjects is given in Figure 2. The median of the distribution is 2.01 years.

3.2 Machine learning

We use machine learning for voxelwise prediction of amyloid positive subjects (PreAD) among cognitively unimpaired subjects. A realistic prevalence for PreAD subjects on middle-age adults is 20% [26]. We use this prevalence to fix the proportion of PreAD on the test set on all machine learning experiments.

Another key parameter of the analysis is the temporal distance (Δt) between reference and target images used to compute the Jacobian determinant maps. In Figure 3 we report the performance of the classifier as a function of minimal Δt values in the test set. It is observed that even though we normalize each Jacobian determinant map with respect to the Δt parameter, the preclinical signature is within the detection range when visits are at least 2.5 years apart. In the case in which $\Delta t > 2.5$ years, the performance of the classifier based on structural changes is much better than a classifier trained on individual images as reported in our previous cross-sectional study that reports an AUC = 0.76 [13]. When using Jacobian determinant maps with smaller temporal distance ($\Delta t < 2.5$ years), the mean performance is worse than the cross-sectional analysis, probably due to the low signal to noise ratio between the changes due to normal brain ageing and the changes due to amyloid positivity [13].

The optimal temporal span in terms of AUC and savings between data acquisitions is $\Delta t > 2.5$ years. The number of subjects with follow-up visits between $2.5 < \Delta t < 3.5$ years from baseline is reduced to 18 Ctrls, 10 PreAD and 39 MCI/AD subjects with 25, 16 and 52 Jacobian determinant maps respectively. In what follows, throughout the paper, we use only Jacobian determinant maps within the optimal temporal span ($2.5 < \Delta t < 3.5$ years).

Receiver operating characteristic curve (ROC) and precision-recall (PR) curves of the classifier are shown in Figure 4. A savings heatmap that responds to Eq. 1 is overlaid on the PR curve, while the mean and standard deviation of the model performance are plotted against the random classifier on the ROC curve.

The impact of different number of features used to train our multivariate algorithm is presented in Table 3, evaluated on our dataset which is imbalanced (36% of preclinical subjects). Note that the prevalence of preclinical subjects on the test set is forced to 20% using permutations. When using a low number of features, the model underrepresents the preclinical signature, not being able to capture all data heterogeneity. In contrast, when using a high number of features, the model is not able to generalize results to unseen Jacobian determinant maps, overfitting the development set. Hence, the best results are obtained using a moderate number of features that are able to both represent the preclinical signature and still generalize well to the test set.

An optimal compromised solution between several metrics is to design our model using a 0.5% of the total Jacobian features. In this case, after the 100 iterations of the nested cross-validation framework, a heatmap of selected features is shown in Figure 5. As expected, the top selected features correspond to typical regions affected by AD pathology like the caudates, fusiform or parahippocampal gyrus, presenting high overlap with the statistical analysis presented in the next section. This result shows that a machine learning classifier trained on changes in specific brain regions has the capacity to predict the presence of early amyloid pathology in asymptomatic individuals as measured by MRI.

3.3. Statistical analysis:

In parallel to the machine learning classification model, we carried out a voxel-wise statistical analysis using the full dataset of Jacobian determinant maps to identify the regions of volumetric change that are statistically significant between the different categories Ctrl, PreAD and AD/MCI (Figure 6).

3.3.1 Preclinical signature

Stable PreAD individuals show significantly higher grey-matter (GM) atrophy in the parahippocampal and fusiform gyri as compared to amyloid negative cognitively unimpaired subjects, as shown on the left hand

side of Figure 6. Apparent mild GM increments are detected in the caudate heads, probably as a surrogate effect of ventricular expansion.

Furthermore, comparison of longitudinal volumetric changes between amyloid negative cognitively unimpaired subjects and stable symptomatic ones (amyloid positive MCI or AD subjects) reveals the well-known AD signature involving temporo-parietal and posterior cingulate areas, as well as most of the basal ganglia [27], as shown on the right hand side of Figure 6. Of note, apparent GM increments are also detected in periventricular areas, including the caudates and medial thalamus.

4. Discussion

The goal of this work was to assess whether brain structural changes captured by subsequent magnetic resonance images can indicate the presence of abnormal amyloid levels in cognitively unimpaired subjects using machine learning-techniques. In addition, we also aimed at characterizing the preclinical signature voxelwise using Jacobian determinant maps as a measure of volumetric rate of change.

A machine learning framework was implemented for the classification of amyloid positive subjects using Jacobian determinant maps as features for classification. The best achieved performance in our longitudinal classifier (AUC: 0.866) significantly improved the performance we previously reported for a cross-sectional classifier (AUC 0.76) [13]. This performance is significantly higher than what was reported in previous works that, on top of using MRI ROI data, built classifiers adding demographics (AUC: 0.63), demographics and genetics (AUC: 0.62-0.66) and demographics, neuropsychology and APOE (AUC: 0.74) [14, 28]. Therefore, there is still room for improvement in the performance of our classifier by adding complementary information such as demographics and genetic risk factors.

The increased performance of our classifier may be accounted for two factors. On the one hand and unlike similar previously reported classifiers, we used voxelwise data as features. Coupled with an efficient feature selection strategy, this allowed the classifier to select the most discriminant brain regions, independent of *a-priori* cortical parcellations. These regions mostly included AD-related areas in the medial and inferior temporal lobe, as well as the lateral ventricles which can be considered as the *preclinical AD signature*. On the other hand, we used subsequent images that correspond to the same individuals, thus eliminating an important percentage of the between-subject variability present in cross-sectional setups.

In this regard, we observed that our classifier works significantly better only when the pairs of MRI scans that are used for test and training sets are acquired more than 2.5 years apart. This time period is likely related to the protracted evolution of the neuroanatomical changes in preclinical AD stages. At more advanced stages of the disease, more rapid evolution of brain structural changes is expected, and thus, the benefits of a longitudinal classifier would potentially be evident with shorter time intervals. It remains to be explored how these promising results would be affected by the use of different scanners.

The predictive capacity achieved by this classifier does not place this method as substitute of gold-standard tests to detect amyloid abnormalities. Still, if used for triaging of subjects, e.g. clinical trial recruitment, we demonstrated that it could allow significant savings in terms of the number of costly gold-standard tests that would have to be performed to detect a fixed number of amyloid positive, cognitively healthy subjects. Used in this way, in a cognitively unimpaired population with a prevalence of amyloid positivity of 20%, the accuracy of the longitudinal classifier would allow a reduction of up to 55% of unnecessary PET or CSF tests, which translates to a 40% reduction of the total cost, according to the savings model we previously proposed [13]. Nevertheless, in a clinical trial recruitment setting, it can be more advantageous instead to optimize the sensitivity of the classifier to maximize the number of detected at-risk individuals, at the cost of a slightly poorer specificity which might decrease these cost savings.

Due to the limited sample size for training and the large inter-subject variability of cerebral morphology, we use a simple but effective model for prediction of amyloid positivity. Our method is fully automatic from feature extraction and signature learning to classification. However, the presence of high-dimensional and low informative features together with the overlap between normal aging and AD processes in the brain, reduces the overall precision of the system. To account for that, future efforts will need larger longitudinal datasets and many initiatives are contributing to achieve this [14,29].

On top of this, given the limited sample size and the large amount of features used for classification (voxels), we might have incurred in an overfitting of the existing data, potentially resulting in an overestimation of the capacity of the classifier. Therefore, our results need to be validated on independent datasets, but the scarcity of longitudinal MRI datasets with CSF biomarker levels has prevented us to conduct such validation in this work. Still, in our previous ROI-based study, we successfully validated a very similar classifier with two independent datasets without a major loss of the classifier's performance [13].

To further characterize the *preclinical AD signature*, a statistical analysis was conducted and we report longitudinal morphological changes in cognitively unimpaired subjects with abnormal amyloid CSF levels. This preclinical AD signature comprises atrophy of the parahippocampal and fusiform gyri and expansion of the lateral ventricles. This pattern is in line with previous reports of longitudinal volumetric changes

associated to the presence of abnormal amyloid levels from ADNI participants that have been replicated in an independent cohort [30]. On the other hand, expansion of the caudate heads falls beyond this known pattern. Being in the proximity of the lateral ventricles, it may be questioned whether the detected increase in the volume of the caudates is an actual feature associated to preclinical AD stages or an artifact of the processing methodology to detect volumetric changes. By smoothing spatially continuous Jacobian determinant maps, it could be considered that the observed increase in caudate volumes could be a side effect of the 'spillover' of the Jacobian determinant maps due to the expansion of the ventricles. To address this question, we performed a *post-hoc* analysis of the caudate volumes between the Ctrls and PreAD groups, but using the longitudinal Freesurfer pipeline to compute change in caudate volumes. Since the subcortical segmentation implemented in Freesurfer uses an ROI-approach based on a probabilistic atlas [31], it can be considered to be virtually free from the potential 'spillover' effect of continuous Jacobian determinant maps. Results show that the changes in caudate volumes is not significantly different between Ctrls and PreAD individuals ($p > 0.3$) and, thus, it can be concluded that the observed caudate head expansion is artifactual and secondary to ventricular expansion. Still, this signal might contribute to the detection of the presence of amyloid-burden in cognitively unimpaired individuals.

This study has some limitations. Even though data comes from a heterogeneous sample with different sites, and MRI scanners, the MRI acquisition was harmonized according the ADNI protocol. Therefore, the performance of our method when applied to MRI samples using different acquisition protocols may deviate from what is here reported. Actually, the ultimate validation of the generalizability of the results here reported can only be accomplished by applying the method here developed to an independent sample. In our previous work, the performance of a similar cross-sectional classifier was kept stable when derived and validated in two independent cohorts. Therefore, it can be expected the same behaviour in this longitudinal extension of the classifier. Finally, we used CSF amyloid as the gold-standard for amyloid positivity and not PET imaging. It could be argued that the performance of the classifier could be sensitive the the selection of the gold-standard method. However, the agreement between CSF and PET

determinations of amyloid is very high, particularly in the intermediate ranges where thresholds for positivity typically lie.

To sum up, we here presented a machine learning framework used to predict the presence of amyloid abnormalities in cognitively unimpaired individuals with a moderate-to-high accuracy (AUC: 0.857) when MRI scans acquired 2.5 years apart are available. This performance translates to improvements of up to 55% in the number of necessary CSF/PET tests and a reduction of 40% of the costs to detect a fixed number of amyloid positive individuals. This performance is expected to still have room for improvement by including demographic, genetic and cognitive data to the classifier. We further compare the features used by the classifier with the characteristic pattern of longitudinal morphological changes in preclinical AD that is expressed in typical AD-related regions, uncovering areas that appear to be specific to the preclinical AD stage.

5. Conclusions

In this study, we used longitudinal structural brain MRI scans to predict the presence of amyloid pathology in cognitively unimpaired individuals and unveil the preclinical AD signature. We applied machine learning techniques on Jacobian determinant maps coding longitudinal volumetric changes at the voxel level. This allowed the classifier to significantly improve its performance (AUC = 0.857) with respect to previous cross-sectional ROI-based approximations. Areas showing the most discriminant capacity included medial, inferior and lateral temporal regions, along with the ventricles and caudate heads. The volumetric changes in these areas are in line of those observed in symptomatic stages, but are expressed to a lower extent. Even though the performance of the classifier does not allow for it to substitute gold-standard methods to determine the presence of amyloid pathology, its use as triaging tool would lead to significant reductions of 55% of unnecessary gold-standard tests and of 40% of the cost to detect a fixed number of cognitively healthy individuals in preclinical AD stages. High overlap by the features used by the classifier and the preclinical AD signature is found, characterized by parahippocampal and fusiform gyri atrophy and expansion of the ventricles. To sum up, machine learning over brain longitudinal MRI data can represent a valuable tool for the implementation of secondary prevention trials.

Abbreviations

Ctrls: control subjects; PreAD: preclinical alzheimer's Disease; MCI: mild cognitive impairment; AD: Alzheimer's Disease; MRI: magnetic resonance imaging; CSF: cerebrospinal fluid; PET: positron emission tomography; Aβ: amyloid-beta; SPM: statistical parametric mapping; MNI: Montreal Neurological Institute; ROI: region of interest; AUC: area under the curve; CV: cross-validation; CI: confidence interval; GM: grey-matter

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu), which is easily available for download from the Laboratory of Neuroimaging (LONI) website to the research public.

Competing interests

None

Funding

This work has been partially supported by the project MALEGRA TEC2016-75976-R financed by the Spanish Ministerio de Economía y Competitividad and the European Regional Development Fund (ERDF). Adrià Casamitjana is supported by the Spanish *Ministerio de Educación, Cultura y Deporte* FPU Research Fellowship. Juan D. Gispert holds a '*Ramón y Cajal*' fellowship (RYC-2013-13054).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan

Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Author's contribution

PP, CF and JDG contributed to the conception and design of the study. AC, MA and VV conducted the statistical and machine learning analysis. PP and AC prepared the manuscript and CF, MA, GO, RC, JLM, VV and JDG revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at https:// adni.loni.usc.edu/wp-content/uploads/how_to_apply/ ADNI_Acknowledgement_List.pdf.

References

[1] Gregory, Sarah, et al. Research participants as collaborators: Background, experience and policies from the PREVENT Dementia and EPAD programmes. *Dementia* 17.8 (2018): 1045-1054.

[2] Jansen WJ, Ossenkoppele R, Knol DL, Tijms BM, Scheltens P, Verhey FRJ, et al. Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. JAMA 2015;313:1924–38. doi:10.1001/jama.2015.4668.

[3] Dubois, Bruno, et al. Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia* 12.3 (2016): 292-323.

[4] Frisoni, Giovanni B., et al. The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology* 6.2 (2010): 67.

[5] Dickerson, Bradford C., et al. MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. *Neurobiology of aging* 22.5 (2001): 747-754.

[6] Albert, Marilyn S., et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia* 7.3 (2011): 270-279.

[7] Fennema- Notestine, Christine, et al. Structural MRI biomarkers for preclinical and mild Alzheimer's disease. *Human brain mapping* 30.10 (2009): 3238-3253.

[8] Killiany, R. J., et al. MRI measures of entorhinal cortex vs hippocampus in preclinical AD. *Neurology* 58.8 (2002): 1188-1196.

[9] ten Kate, Mara, et al. Secondary prevention of Alzheimer's dementia: neuroimaging contributions. *Alzheimer's research & therapy* 10.1 (2018): 112.

[10] Falcon, Carles, et al. Longitudinal structural cerebral changes related to core CSF biomarkers in preclinical Alzheimer's disease: A study of two independent datasets. *NeuroImage: Clinical* 19 (2018): 190-201.

[11] Dickerson, B. C., et al. Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults. *Neurology* 76.16 (2011): 1395-1402.

[12] Rathore, Saima, et al. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage*155 (2017): 530-548.

[13] Casamitjana, Adrià, et al. MRI-Based Screening of Preclinical Alzheimer's Disease for Prevention Clinical Trials. *Journal of Alzheimer's Disease* Preprint (2018): 1-14.

[14] ten Kate, Mara, et al. MRI predictors of amyloid pathology: results from the EMIF-AD Multimodal Biomarker Discovery study. *Alzheimer's research & therapy* 10.1 (2018): 100.

[15] ADNI | Alzheimer's Disease Neuroimaging Initiative n.d. http://adni.loni.usc.edu/ . Accessed July 13, 2017.

[16] Jack, Clifford R., et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia* 14.4 (2018): 535-562.

[17] Shaw, Leslie M., et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of neurology* 65.4 (2009): 403-413.

[18] Jack CR, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, et al. Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on

diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:257–62. doi:10.1016/j.jalz.2011.03.004.

[19] SPM - Statistical Parametric Mapping n.d. http://www.fil.ion.ucl.ac.uk/spm/ Accessed July 13, 2017.

[20] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. Neuroimage 2002;15:273–89. doi:10.1006/nimg.2001.0978.

[21] Dukart, Juergen, et al. Age correction in dementia–matching to a healthy brain. *PloS one* 6.7 (2011): e22193.

[22] Saeys, Yvan, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in bioinformatics. *bioinformatics* 23.19 (2007): 2507-2517.

[23] Le Cessie, Saskia, and Johannes C. Van Houwelingen. Ridge estimators in logistic regression. *Applied statistics* (1992): 191-201.

[24] Varoquaux, Gaël, et al. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *NeuroImage* 145 (2017): 166-179.

[25] Pedregosa, Fabian, et al. Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12.Oct (2011): 2825-2830.

[26] Jansen, Willemijn J., et al. Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. Jama 313.19 (2015): 1924-1938.

[27] Neurol, Semin. Neuroimaging biomarkers of neurodegenerative diseases and dementia. *Semin Neurol* 33.4 (2013): 386-416.

[28] Ansart, Manon, et al. Prediction of amyloidosis from neuropsychological and MRI data for cost effective inclusion of pre-symptomatic subjects in clinical trials. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, Cham, 2017. 357-364.

[29] Molinuevo, José Luis, et al. The ALFA project: a research platform to identify early pathophysiological features of Alzheimer's disease. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* 2.2 (2016): 82-92.

[30] Falcon, Carles, et al. Longitudinal structural cerebral changes related to core CSF biomarkers in preclinical Alzheimer's disease: A study of two independent datasets. *NeuroImage: Clinical* 19 (2018): 190-201.

[31] Fischl, Bruce, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33.3 (2002): 341-355.

[32] Gispert, Juan Domingo, et al. Nonlinear cerebral atrophy patterns across the Alzheimer's disease continuum: impact of APOE4 genotype. *Neurobiology of aging* 36.10 (2015): 2687-2701.

[33] Ashburner, John. "A fast diffeomorphic image registration algorithm." *Neuroimage* 38.1 (2007): 95-113.

[34] Evans, Alan C., et al. "3D statistical neuroanatomical models from 305 MRI volumes." *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.*. IEEE, 1993.

FIGURE LEGENDS

Figure 1: Workflow of the optimization and evaluation of the classification method. The performance of the final classifier is evaluated on a fresh test set that has not been used for training.

Figure 2: Distribution of the interval Δt between reference and follow-up visits across the whole dataset.

Figure 3. AUC and savings (blue, green) reported using Jacobian determinant maps with different time intervals (Δt) between reference and target and a fixed prevalence of 20% amyloid positive subjects on the test set. To compute savings we used optimal precision and recall values plotted in dashed orange and red lines, respectively using the cost function defined in Eq. 1.

Figure 4. ROC and PR curves for Jacobian determinant maps with time spans in the range $2.5 < \Delta t < 3.5$ years using 0.5% of the features. On the left, the ROC curve is averaged across different development/test splits: the mean curve (blue) with the standard deviation (gray) and the curve of a random classifier (red). On the right, the PR curve of the classifier (blue) is overlaid on a savings heatmap (Eq. 1). Black lines indicate points of equal savings.

Figure 5. Normalized feature maps of the 0.5% of features selected during the 100 different splits of the development/test sets, representing the frequency of selection of each feature. Those features have optimal capacity to detect the presence of early amyloid pathology in asymptomatic individuals.

Figure 6: Statistical maps for group comparison between Ctrl and PreAD (PreAD signature) and Ctrl and MCI/AD (AD signature) subjects. Statistical significance was set to uncorrected p-value < 0.005 and minimum spatial extent k > 100.

Number of visits	N
2	295
3	63
4	27
5	15
6	3
Total	403 subjects 841 Jacobian maps

Table 1: Distribution of the number of 3D-T1 MRI acquisitions per subject.

Category	**Ctrl (Aβ-)**	**PreAD (Aβ+)**	**MCI (Aβ+)**	**AD (Aβ+)**	**MCI/AD (Aβ+)**
Number of subjects	79	50	196	78	274
Age (years) at baseline (mean; std)	73.97 (5.97)	76.04 (6.25)	73.55 (6.55)	75.44 (7.37)	74.1 (6.85)
Sex (F/M)	37/42	21/29	79/117	33/45	112/162
Follow-up (years) period (mean; std)	2.48 (1.38)	2.32 (1.32)	2.2 (1.09)	1.4 (0.46)	1.97 (1.02)

Table 2. Dataset demographics at baseline.

#features (%)	AUC (95% CI)	Accuracy (95% CI)	Precision (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Fscore (95% CI)
6 (0.001)	0.782 (0.5-0.99)	0.568 (0.2-0.8)	0.334 (0.13-0.5)	0.907 (0.33-1.0)	0.483 (0-0.75)	0.477 (0.19-0.67)
65 (0.01)	0.809 (0.6-0.97)	0.631 (0.27-0.73)	0.347 (0.18-0.43)	0.91 (0.33-1.0)	0562 (0.12-0.75)	0.499 (0.28-0.6)
653 (0.1)	0.846 (0.67-1.0)	0.651 (0.53-0.8)	0.368 (0.27-0.5)	0.97 (0.67-1.0)	0.571 (0.42-0.75)	0.531 (0.38-0.67)
1633 (0.25)	0.857 (0.72-1.0)	0.648 (0.46-0.8)	**0.37 (0.27-0.5)**	**0.98 (0.67-1.0)**	0.534 (0.33-0.75)	**0.533 (0.43-0.67)**
3266 (0.5)	0.857 (0.71-0.97)	0.641 (0.46-0.8)	0.364 (0.26-0.5)	0.973 (0.67-1.0)	0.558 (0.33-0.75)	0.526 (0.38-0.67)
6532 (1)	**0.866 (0.72-0.97)**	**0.653 (0.53-0.8)**	0.363 (0.25-0.5)	0.933 (0.67-1.0)	**0.583 (0.42-0.75)**	0.521 (0.36-0.67)
13064 (2)	0.855 (0.64-1.0)	0.645 (0.46-0.8)	0.359 (0.2-0.5)	0.917 (0.67-1.0)	0.577 (0.33-0.75)	0.512 (0.31-0.67)
32661 (5)	0.802 (0.49-1.0)	0.57 (0.4-0.77)	0.301 (0.14-0.47)	0.837 (0.33-1.0)	0.503 (0.33-0.71)	0.442 (0.33-0.71)
65323 (10)	0.77 (0.4-1.0)	0.573 (0.4-0.77)	0.298 (0.14-0.47)	0.813 (0.33-1.0)	0.512 (0.33-0.75)	0.434 (0.33-0.75)

Table 3. Performance of the system using a different number of features evaluated on the interval $\Delta t > 2.5$ years

Figure 1

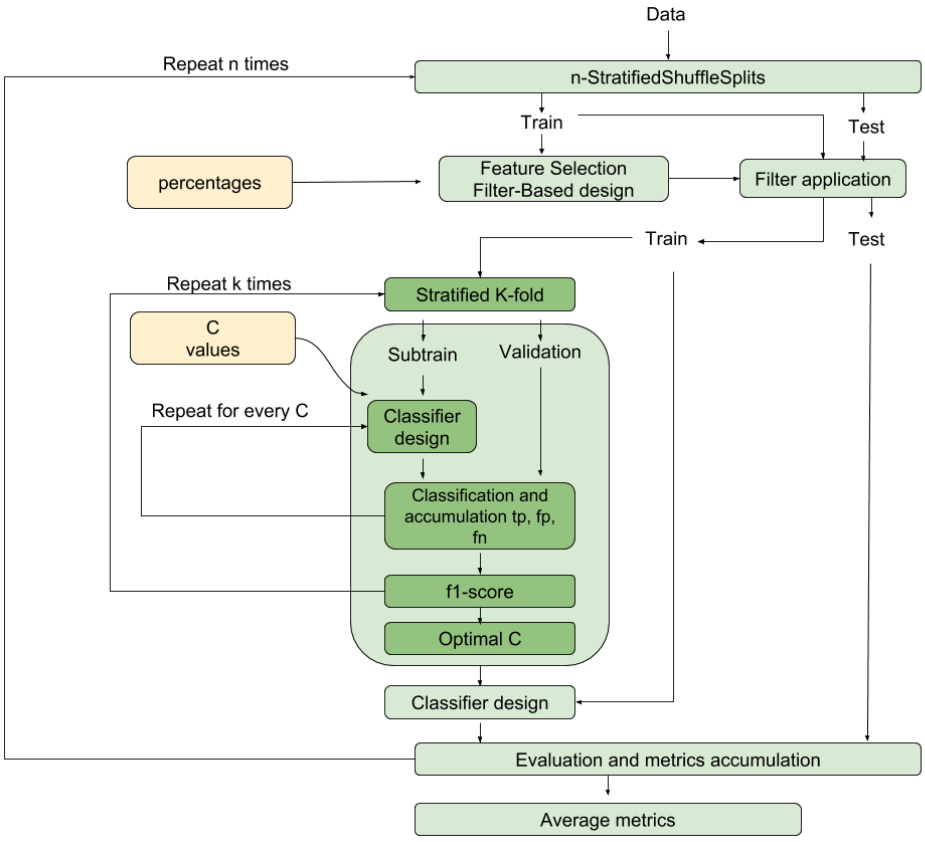


Figure 2

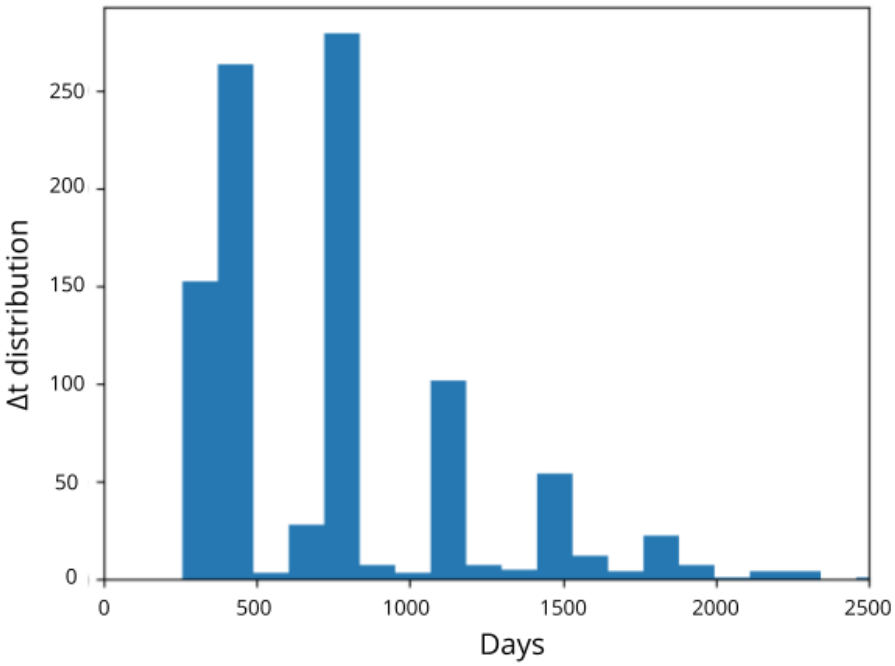


Figure 3

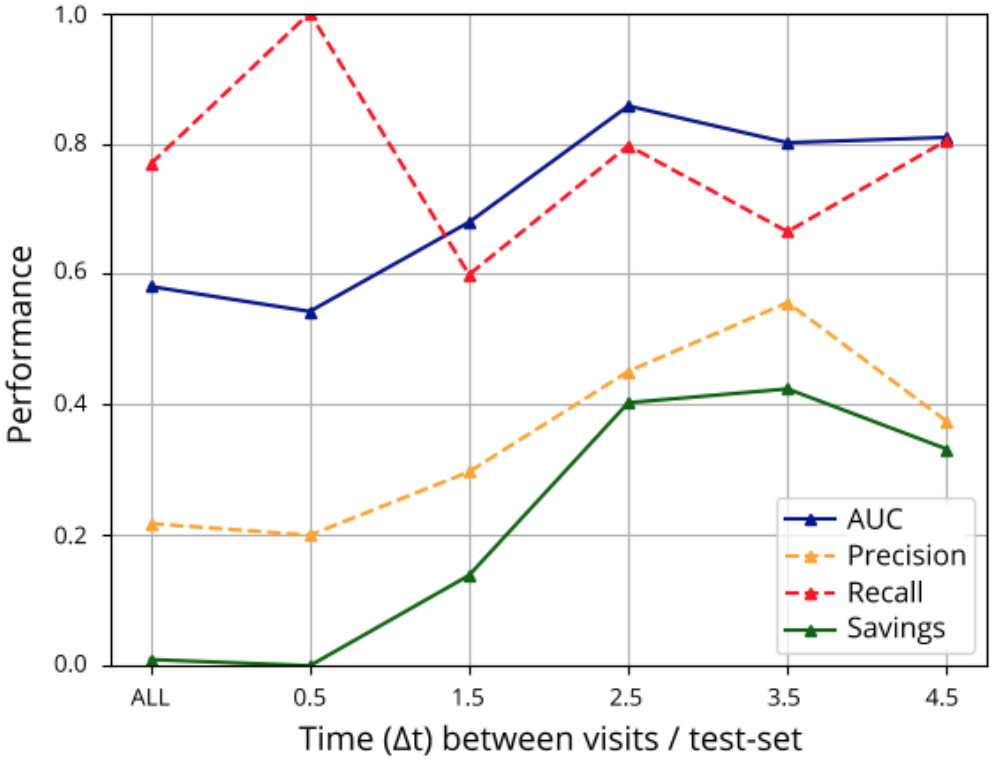
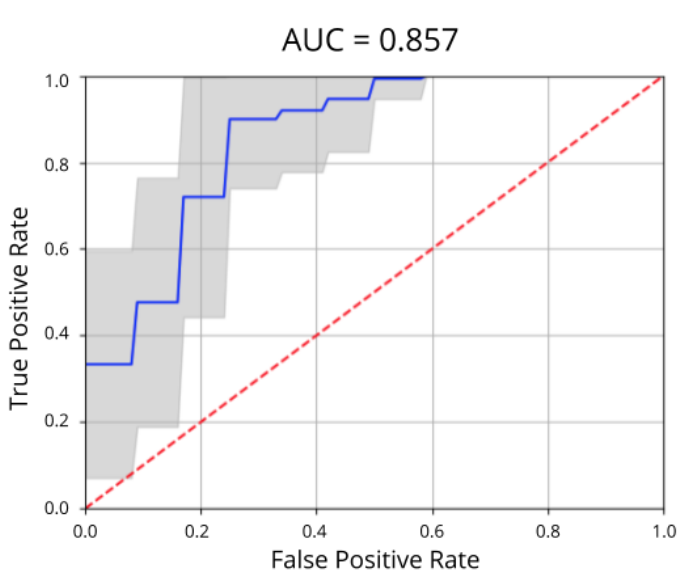
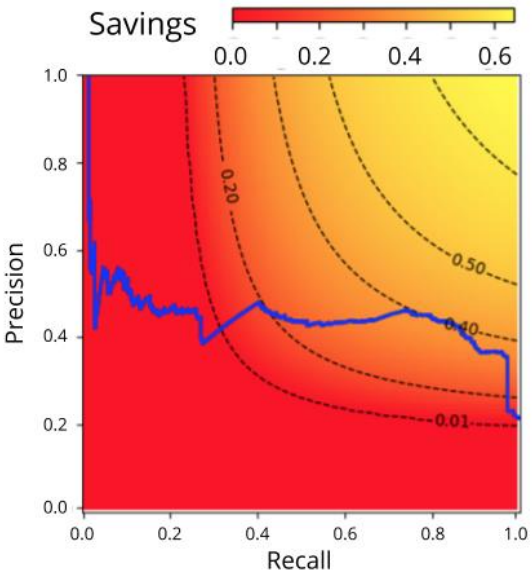


Figure 4



(a) ROC curve



(b) Precision-Recall curve

Figure 5

Click here to access/download;Figure;Petrone_Fig5.pdf

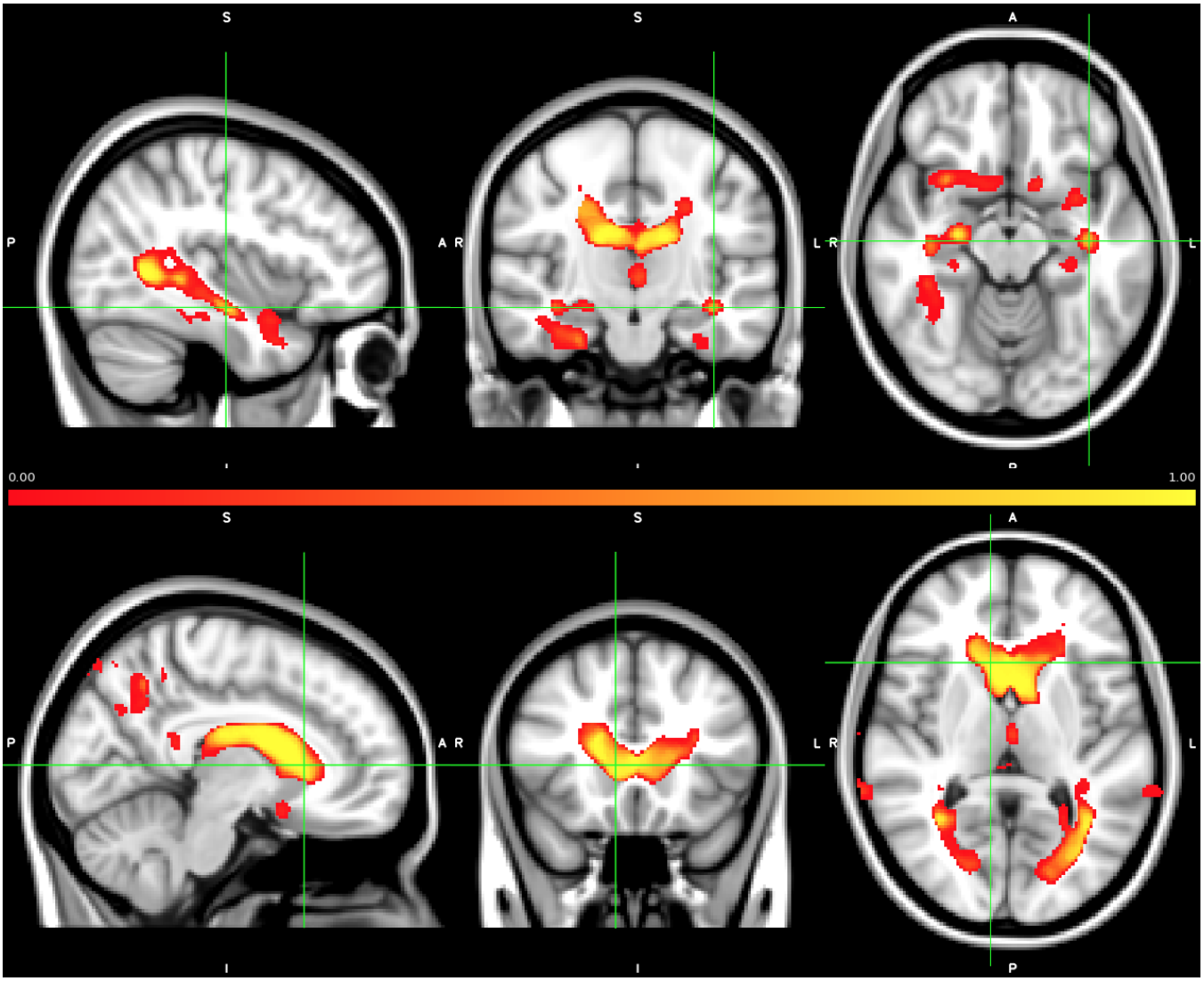


Figure 6